Codebook

This guide can be used to reproduce the data analyses reported in "Mwar: A Bayesian Estimator of Manager Value," now in press in the *Journal of Sports Analytics*. The necessary data and code are posted on supporting materials page. This code book catalogs and describes the data and code files. All the data are in .csv format. The codes consist of either Stata .do or Python .ipynb files; they are amply annotated to enable translation into other programming languages.

A. Data files

df_1.manager_gamel_level.csv

game-level manager record

variable name	description	note
retroID	manager retroID	
game_date	game date	
game_id	retrsheet game id	
teamIDretro	retrosheet team name	
opponent	retrosheet opponent	
team_runs	team runs	
opponent_runs	opponent runs	
result	game result	
home_away	home (H) or away (A)	
name_last	manager last name	
name_first	manager first name	
yearID	season	
teamID	Lahman teamID	

df_2.team_warpg.csv

team WAR per game

variable name	description	note
teamID	Lahman teamID	
yearID	season	
warpg	WAR per game	
z_wpg	season standardized WPG	

df_3.mwar_bootstrap_feeder.csv

data for parametric bootstrap null

variable name	description	note
mgrID	manager retroID	
name_first	manager last name	
name_last	manager first name	
yearID	season	
teamID	team	
w	wins	
l	losses	
games	season games	
wp	winning percentage	
warpg	team WAR per game	
z_wpg	season standardized WPG	
mgr_n	unique mgr numeric identifier	
tg	career games managed	vestige of data cutoff

df_4.observed_mwar_mle.csv

manager observed MLE data

variable name	description	note
mgrID	manager retroID	
mgr_n	unique mgr numeric identifier	
xb	predicted manager effect logit scale	

df_5.mwar_bs_en.csv

bootstrap simulated realisitic null

variable name	description	note
rep	replicate number	
mgr_n	unique mgr numeric identifier	
xb_en	replicate predicted manager effect logit scale	
mgrID	manager retroID	
xb_mu	manager sim mean	mean of replicate scores
xb_se	manager sim SE	SD of replicate scores
z_sim	manager z	(rep xb_en - xb_mu)/xb_se

df_6_hb_mcmc_replicates

Hierarchical Bayes MCMC draws

variable name	description	note
draw	MCMC draw no.	
mgrID	manager retroID	
u	manager theta logit scale	
prob	manager theta prob scale	
w162	manager theta mw162	

df_7.mcmc_replicates_plus_mle_data.csv

Hierarchical Bayes MCMC draws plus plus MLE data scores

variable name	description	note
mgrID	manager retroID	
MAP_w162	posterior MAP mw162	same value for all manager rows
HDI_lower_w162	lower bound 0.95 HDI	same value for all manager rows
HDI_upper_w162	upper bound 0.95 HDI	same value for all manager rows
draw	MCMC draw no.	
u	MCMC draw no. theta	
data_162	MLE data score wp	same value for all manager rows
data_162se	MLE data SE wins/162	same value for all manager rows
mw162	MLE data score	same value for all manager rows

df_8.mwar_MAP_95hdis

Hierarchical Bayes MAPs and 0.95 HDIs

variable name	description	note
mgrid	manager retroID	
map_w162	posterior MAP mw162	
hdi_lower_w162	lower bound 0.95 HDI	
hdi_upper_w162	upper bound 0.95 HDI	
hdi_width	0.95 posterior pass width	
mean_w162	mean theta in mw162	
median_w162	median theta in mw162	
sd_w162	sd theta wm162	
ci_lower_w162	ci lower theta mw162	
ci_upper_w162	ci uppper theta mw162	
ci_width	ci width theta mw162	
n_draws	total number MCMC draws	

df_9.team_recs_plus_wpg.csv

season team records plus WAR per game scores

variable name	description	note
yearID	season	
w	wins	
l	losses	
teamID	Lahman teamID	
warpg	team WAR per game	
z_wpg	season standardized WPG	
games	season games played	
wp	winning percentage	

df_10.2025_mgr_patchcsv.csv

supplemental patch for 2025 manager records

variable name	description	note
name	manager name	
mgrID	retroID	
team	team managed	
w	wins	
l	losses	
name_first	first name	
name_last	last name	
teamID	Lahman team ID	
yearID	season	

df_11.manager_team_games.csv

number of managed team games vs total team games per season

variable name	description	note
mgrID	manager retroID	
name_first	first name	
name_last	last name	
teamID	Lahman teamID	
yearID	season	
gm	games managed	
gp	team games played	
retroID	manager retroID	

df_12.manager_name_id_match.csv

manager name and manager ID key

variable name	description	note
retroID	manager retroID	
name_last	first name	
name_first	last name	
mgrID	manager retroID	

df_13.mgr_season_by_season_plus_team_wpg.csv

manager record by season along with team WAR per games

variable name	description	note
mgrID	manager retroID	
name_first	first name	
name_last	last name	
yearID	seaspm	
teamID	Lahman teamID	
w	wins	
l	losses	
games	season games managed	
wp	winning percentage	
warpg	team WAR per game	
z_wpg	season standardized WPG	
mgr_n	unique mgr numeric identifier	
tg	career total managed	same value for all manager rows

df_14.season_by_season_mw162_incl_game_count.csv

record of updated manager mw162 per season

variable name	description	note
name_last	first name	
name_first	last name	
yearID	seaspm	
teamID	Lahman teamID	
mgrID	manager retroID	
w	wins	
l	losses	
games	games managed season	
u_map	season-running MAP logit scale	
u_sd	season-running SD logit scale	

tau_fixed	HB tau	model-fixed value for all rows
beta0_fixed	HB constant	model-fixed value for all rows
beta_wpg_fixed	HB beta WAR per game	model-fixed value for all rows
winsHB	season-running MAP mw162	
winsHB_se	season-running mw162 SE	
pHB_se	season-running theta SE prob scale	
cum_games	running game total	

df_15.season_by_season_theta_i_progression_logit_space.csv

record of updated manager estimated MAP in logit space

variable name	**description**	**note**
mgrID	manager retroID	
u_map	season-running MAP logit scale	
u_sd	season-running SD logit scale	
as_of_year	season	
tau_fixed	tau	model-fixed value for all rows
beta0_fixed	HB constant	model-fixed value for all rows
beta_wpg_fixed	HB beta WAR per game	model-fixed value for all rows

df_16.map_plus_data.csv

manager MAP mw162 plus MLE data scores

variable name	**description**	**note**
mgrID	manager retroID	
map_w162	MAP mw162	
hdi_lower_w162	posterior MAP mw162	
hdi_upper_w162	lower bound 0.95 HDI	
hdi_width	upper bound 0.95 HDI	
mean_w162	0.95 posterior pass width	
median_w162	mean theta in mw162	
sd_w162	median theta in mw162	
ci_lower_w162	sd theta wm162	
ci_upper_w162	ci lower theta mw162	
ci_width	ci uppper theta mw162	
n_draws	ci width theta mw162	
mgr_n	total number MCMC draws	
xb	manager MLE impact logit scale	
xb_se	manager MLE impact SE logit scale	
z_obs	MLE z	
se_pr	manager MLE impact SE prob scale	
pr	manager MLE impact prob scale	
data_162	manager MLE wins 162	
data_162se	manager MLE wins 162 SE	

B. Codes

Code No.	description	programming language
1	create observed manager impacts and bootstrap feeder	stata
2	generate bootstrap realistic null	python
2a	code bootstrap realistic null	stata
3	generate and analyze pivotal zs, including global FDR	stata
4	hierarchical Bayes model MCMC draws	python
5	ROPE analysis	python
6	Bayesian updating illustrator	stata
7	mw162 density plot	stata
8	analysis of mw162 ± 2 populaion frequency	python
9	customizable asymmetric loss function	python
10	generation of mw162*p* scores, plus density distribution	stata
11	Plotting of learning curve individual managers	python
12	Aggregate learning curve analysis	stata
13	Analysis of mWAR explanator powr	stata